

Ελένη Μαμμά

Ειδησεογραφικά Πρότυπα Κωδικοποίησης Δεδομένων: Μοντέλο Υλοποίησης του Προτύπου NewsML

Ελένη Μαμμά
Υπουργείο Εξωτερικών
(e_mamma@mfa.gr)

Περίληψη

Σε μία εποχή η οποία χαρακτηρίζεται από ραγδαίες αλλαγές και εξελίξεις στην κοινωνία της πληροφόρησης, η εκδοτική και η δημοσιογραφική κοινότητα, στην προσπάθειά τους να οργανώνουν και να διαχειρίζονται τις πληροφορίες που προέρχονται από τον ειδησεογραφικό χώρο, χρησιμοποιούν ευρέως αναγνωρισμένα πρότυπα μεταδεδομένων. Σκοπός τους η πραγματοποίηση διαδικασιών, όπως η περιγραφή, η κωδικοποίηση, η τεκμηρίωση και η ανταλλαγή δεδομένων μεταξύ μεγάλων φορέων. Στην εκπλήρωση του συγκεκριμένου σκοπού συμβάλλουν πρότυπα όπως, τα NewsML, Nitf, Prism, XMLNews, IPTC 7901, IIM, RSS, ICE, τα οποία προέκυψαν μέσα από μακρόχρονες ερευνητικές προσπάθειες και συνεργατικά σχήματα μεγάλων οργανισμών όπως είναι οι διεθνείς ενώσεις Τύπου και Τηλεπικοινωνιών, οι δημοσιογραφικοί οίκοι, γνωστά πρακτορεία ειδήσεων, κ.ά.

Στόχος της ακόλουθης μελέτης είναι κατ' αρχήν ο εντοπισμός και η επισκόπηση των προαναφερθέντων προτύπων και ακολούθως η κωδικοποίηση ειδησεογραφικών αντικειμένων με την εφαρμογή του προτύπου NewsML. Θα δοθούν τα κύρια χαρακτηριστικά τους γνωρίσματα καθώς και συγκεκριμένες ομοιότητες & διαφορές τους. Επιπροσθέτως, θα ακολουθήσει η κωδικοποίηση ενός ειδησεογραφικού αντικειμένου και η περιγραφή του με βάση το διεθνώς αναγνωρισμένο και χρησιμοποιούμενο πρότυπο περιγραφής ειδησεογραφικών μεταδεδομένων, το επονομαζόμενο NewsML. Ιδιαίτερα η κωδικοποίηση αποτελεί ένα δημιουργικό και ενδιαφέρον κομμάτι στην εφαρμογή, λόγω των σύγχρονων τεχνολογιών, οι οποίες απαιτούνται για την υλοποίηση της διαδικασίας αυτής. Τέλος, θα καταλήξουμε σε ορισμένα ουσιαστικά συμπεράσματα, τα οποία θα αποτελέσουν σημεία για ένα γόνιμο επιστημονικό προβληματισμό όσον αφορά τη χρήση των ειδησεογραφικών προτύπων και την εφαρμογή τους για τη διαχείριση πληροφοριών, όπως είναι τα ειδησεογραφικά αντικείμενα, και την παροχή εξατομικευμένων υπηρεσιών στους τελικούς χρήστες.

Κύριο μέλημα της συγκεκριμένης εργασίας είναι αφενός, να γνωστοποιήσει την ύπαρξη, τη χρησιμότητα καθώς και την αναγκαιότητα αυτών των προτύπων στους κόλπους της Βιβλιοθηκονομίας και της εν γένει Κοινωνίας της Πληροφόρησης, και αφετέρου να αποτελέσει ένα ουσιαστικό και λειτουργικό βοήθημα στο χώρο της Επιστήμης της Πληροφορίας για τη διαχείριση της γνώσης.

Λέξεις – κλειδιά: Ειδησεογραφικά Πρότυπα, Διαχείριση Δεδομένων, Κωδικοποίηση Ειδησεογραφικών Αντικειμένων, Μεταδεδομένα

Abstract

In the epoch of latest rapid changes and evolutions in information society, the publishing and journalistic communities, in their attempt to organize and manage information, which proceed from the environment of news, use broadly recognized

metadata standards. Their goal is the achievement of processes, such as the description, the encoding, the documentation and the exchange of data between big organizations. For the fulfillment of this certain aim some standards contribute such as *NewsML*, *Nitf*, *Prism*, *XMLNews*, *IPTC 7901*, *IIM*, *RSS*, *ICE*, which have come to light after the long-term research efforts and the cooperative schemes of great organizations. The international unions of Press and Telecommunications, big journalistic houses, well-known news agencies are indicatively some of them.

The main goal of the following article is initiatively the localization and the inspection of the above-mentioned standards, and subsequently the encoding of news objects with the implementation of *NewsML standard*. Their main characteristics and several similarities and dissimilarities will be given. Furthermore, the description and the encoding of a certain news object based on the worldwide recognized and used standard of news metadata, denominated as *NewsML*, will follow. Especially, the encoding constitutes a creative and interesting implemented process, because of the contemporary technologies, which are demanded for its materialization. Finally, we will end up to some interesting conclusions that will lead to a fruitful scientific speculation as far as the usage and the implementation of news standards are concerned. These standards are very important for the management of news information and the provision of personalized services to end-users.

The main solicitude of this study is to make known the presence, the usefulness and the necessity of news standards in the bosom of Library Science and generally of Information Society. Moreover, news standards can be a substantial and functional aid in the area of Information Science, concerning knowledge management.

Keywords: *News Standards, Data Management, News Objects Encoding, Metadata*

Εισαγωγή

«Οι προμηθευτές συνήθιζαν να πιστεύουν ότι η ικανότητα να διατίθενται οι ειδήσεις ήταν ένα κομμάτι – κλειδί της αξίας τους. Αυτό έχει αλλάξει. Έχουμε περάσει από ένα κόσμο όπου ήταν δύσκολο να συλλέξουμε ειδήσεις, σε ένα κόσμο όπου είναι δύσκολο να συλλέξουμε μόνο τις ειδήσεις που επιθυμούμε. Σαν αποτέλεσμα, έπρεπε να σκεφτόμαστε προσεκτικά σχετικά με το τί κάνουμε και πώς προσθέτουμε αξία» (Rabin, 2003, p. 1)

Μία άποψη, η οποία περικλείει τις βασικές έννοιες: ειδήσεις, οργάνωση και διαχείριση αυτών. Η ακόλουθη εισήγηση πραγματοποιείται μέσα από μία προσπάθεια να εντοπισθούν και να αναπτυχθούν ορισμένα ευρέως υιοθετημένα από την κοινότητα της Επιστήμης της Πληροφορίας πρότυπα για την οργάνωση, δόμηση και διαχείριση πληροφοριών και συγκεκριμένα των ειδήσεων.

Στον τομέα του εκδοτικού και πιο ειδικά του δημοσιογραφικού περιβάλλοντος υπάρχουν πρότυπα, όπως τα *NewsML*, *Nitf*, *Prism*, *XMLNews*, κ.ά. Αυτά τα πρότυπα είναι σημαντικά και βασικά εργαλεία για την περιγραφή, την οργάνωση, την τεκμηρίωση, ανταλλαγή και διαχείριση ειδησεογραφικών δεδομένων μεγάλων φορέων, όπως είναι οι εφημερίδες, τα πρακτορεία ειδήσεων, οι εκδοτικοί οίκοι, κλπ.

Κύριος σκοπός της συγκεκριμένης εισήγησης είναι να μας εισάγει στον κόσμο των ποικίλων προτύπων, τα οποία υπάρχουν και εφαρμόζονται στη βιομηχανία ειδήσεων και συγχρόνως να γνωρίσουμε και να κατανοήσουμε τη χρησιμότητα και την αναγκαιότητα της ύπαρξής τους τόσο στο δημοσιογραφικό τομέα, όσο και στον τομέα της εν γένει Κοινωνίας της Πληροφόρησης.

Επισκόπηση Κυριότερων Προτύπων για τη Διαχείριση Ειδήσεων

Τα διάφορα πρότυπα, τα οποία έχουν χρησιμοποιηθεί ή συνεχίζονται να χρησιμοποιούνται από τη βιομηχανία ειδήσεων, θα μπορούσαν να ενταχθούν σε μία κατηγοριοποίηση το καθένα από αυτά ανάλογα με τις ανάγκες που καλύπτουν και διευθετούν με την χρήση τους. Μπορούμε να πούμε ότι, με βάση το συμβατικό τρόπο με τον οποίο οι ειδήσεις μεταδίδονταν μεταξύ των διαφόρων οργανισμών, υπάρχουν διάφορες κατηγορίες και ποικίλα συστατικά μέρη για να εντάξουμε τα συγκεκριμένα πρότυπα (Dumbill, 2000). Έτσι έχουμε:

- ◆ Το **Πρωτόκολλο (Protocol)**: χρησιμοποιείται για να μεταφέρει την πληροφορία μεταξύ των ενδιαφερόμενων μερών πάνω στο μέσο μετάδοσης
- ◆ Ο **Φάκελος (Envelope)**: χρησιμοποιείται για να αναγνωρίζει ένα τμήμα πληροφορίας

<i>Protocols</i>	<i>Envelope</i>	<i>Metadata-Header</i>	<i>Content</i>
ICE			NITF
HTTP	NewsML	PRISM	XML – News-Story
FTP		XML - News	Υπάρχοντα πολυμεσικά μορφότυπα

- ◆ Η **Επικεφαλίδα (Header)**: χρησιμοποιείται για την αναγνώριση των μεταδεδομένων ενός ειδησεογραφικού αντικειμένου
- ◆ Το **Περιεχόμενο (Content)**: χρησιμοποιείται για το πραγματικό (καθαυτό) περιεχόμενο του αντικειμένου (Dumbill 2000)

➤ **NEWSML**

Το **NewsML** (News Markup Language) έχει αναπτυχθεί και επικυρωθεί ως ένα ανοικτό πρότυπο από το Διεθνές Συμβούλιο Τύπου και Τηλεπικοινωνιών (*International Press Telecommunications Council*) για τη δόμηση και διαχείριση ειδήσεων. Το Διεθνές Συμβούλιο Τύπου και Τηλεπικοινωνιών ιδρύθηκε το 1965 με σκοπό τη διαφύλαξη των τηλεπικοινωνιακών συμφερόντων και ενδιαφερόντων του παγκόσμιου τύπου και μία από τις πρωταρχικές δραστηριότητές του είναι η δημιουργία εργαλείων και προτύπων για την ανταλλαγή δεδομένων ειδήσεων.

Το πρότυπο **NewsML** αποτελεί ένα από αυτά τα εργαλεία. Στην ουσία έχει σχεδιαστεί για να παρέχει τη δυνατότητα μεταφοράς πολυμεσικών αντικειμένων (κείμενο, φωτογραφία, ήχος, βίντεο) ανεξαρτήτως της κωδικοποίησης, της γλώσσας, του μορφοτύπου και του επιπέδου μεταφοράς που χρησιμοποιούνται, και σε συνδυασμό με την παραγωγή, τη χρήση και την επαναχρησιμοποίηση των ειδήσεων σε ολόκληρο τον κόσμο (Bide, Mark et al, 2000, p. 12). Είναι βασισμένο στη διαδεδομένη γλώσσα σήμανσης XML και σε άλλα κατάλληλα πρότυπα. Έχει ως σκοπό να εμφανίζει ξανά και να διαχειρίζεται κομμάτια ηλεκτρονικών ειδήσεων, συλλογές τέτοιων κομματιών, τις σχέσεις μεταξύ τους, καθώς και τα σχετικά μεταδεδομένα τους μέσα από τον ευρύτερο κύκλο ζωής τους. Επιτρέπει την παροχή πολλαπλών εμφανίσεων της ίδιας πληροφορίας και χειρίζεται αυθαίρετες αναμίξεις τύπων, μορφοτύπων, γλωσσών και κωδικοποιήσεων. (Goodman, 2002)

Επιτρέπει, επίσης, την εξέλιξη των ειδησεογραφικών αντικειμένων μέσα στο χρόνο. Αν και λειτουργεί ανεξάρτητα από το μέσο που φέρει την πληροφορία, το **NewsML** παρέχει ειδικούς μηχανισμούς για να χειρίζεται και κείμενο. Υποστηρίζει την προέλευση τόσο των μεταδεδομένων όσο και του περιεχομένου των ειδήσεων ως

επιβαλλόμενα. Κύρια δραστηριότητά του είναι η ανταλλαγή των ειδήσεων. Λέγοντας «Ανταλλαγή ειδήσεων» εννοούμε τη διαδικασία κατά την οποία μετακινείται όχι μόνο το βασικό περιεχόμενο των ειδήσεων, αλλά και τα δεδομένα που περιγράφουν το περιεχόμενο σε μία περίληψη (π.χ. μεταδεδομένα), πληροφορίες σχετικά με τον τρόπο χειρισμού των ειδήσεων (π.χ. δεδομένα διαχείρισης ειδήσεων) και τέλος πληροφορίες σχετικά με τη διαδικασία μεταφοράς των ειδήσεων (π.χ. δεδομένα ανταλλαγής). (NewsML 1.2, Guidelines Version 1.00, 2004, p. 1)

Το πρότυπο **NewsML** κυκλοφόρησε στην πρώτη του έκδοση (**NewsML 1.0**) τον Οκτώβριο του 2000, ενώ τον Σεπτέμβριο του 2001 βγήκε η δεύτερη έκδοσή του (**NewsML 1.1**). Η πιο πρόσφατη έκδοση (**NewsML 1.2**) πραγματοποιήθηκε τον Οκτώβριο του 2003 και απέδειξε πως το **NewsML** αποτέλεσε ένα σταθερό πρότυπο μέσα στο περιβάλλον της εκδοτικής κοινωνίας.

➡ **NITF**

Το πρότυπο **Nitf** (News Industry Text Format) αναπτύχθηκε μέσα από τη συνεργασία των δύο σημαντικότερων οργανισμών προτυποποίησης για τη βιομηχανία ειδήσεων οι οποίοι είναι το Διεθνές Συμβούλιο Τύπου και Τηλεπικοινωνιών (*International Press Telecommunications Council*) στις Η.Π.Α. και η Αμερικανική Ένωση Εφημερίδων (*National Association of America*). Αποτελεί ένα δωρεάν σε διάθεση πρότυπο, καλά τεκμηριωμένο και υποστηριζόμενο και ευρέως χρησιμοποιούμενο από τους σημαντικότερους εκδότες εφημερίδων καθώς και πρακτορεία ειδήσεων σε ολόκληρο τον κόσμο. Η καθιέρωση της γλώσσας **SGML** (Standard General Markup Language) ως ένα χρήσιμο και εξελισσόμενο εργαλείο μέσα στους εκδοτικούς κόλπους οδήγησε στη δημιουργία και θεμελίωση ενός νέου προτύπου, του **Nitf**. Η γλώσσα **SGML**, όπως και η διάδοχός της **XML**, επέτρεπε στους εκδότες να επινοήσουν δικό τους λεξιλόγιο για να περιγράφουν μεταδεδομένα και περιεχόμενο. Όταν η γλώσσα σήμανσης **XML** παρουσιάστηκε ως ένα υποσύνολο της **SGML**, τότε το **Nitf** τροποποιήθηκε για να γίνει συμβατό. Το **Nitf** χρησιμοποιεί τη γλώσσα σήμανσης **XML** για να καθορίζει το περιεχόμενο και τη δομή των ειδησεογραφικών άρθρων. Τα μεταδεδομένα, που χρησιμοποιεί, εφαρμόζονται μέσα σε όλο το περιεχόμενο της είδησης και με αυτόν τον τρόπο οι πληροφορίες είναι πολύ περισσότερο αναζητήσιμες και χρήσιμες από ότι οι πληροφορίες που δίνονται σε σελίδες **HTML** (Hypertext Markup Language). Με τη χρησιμοποίηση του **Nitf** οι εκδότες έχουν τη δυνατότητα να προσαρμόσουν την εμφάνιση και την αλληλεπίδραση των εγγράφων σε ένα εύρος και με βάση τις εξατομικευμένες ανάγκες των συνδρομητών τους. Αυτά τα έγγραφα μπορούν να μετατραπούν σε **HTML** ή σε οποιοδήποτε άλλο μορφότυπο επιθυμεί ένας εκδότης. Το πρότυπο **Nitf** έχει κυκλοφορήσει σε τρεις εκδόσεις. Η πρώτη έκδοση (**Nitf 2.5**) βγήκε το Σεπτέμβριο του 2000, η δεύτερη (**Nitf 3.0**) τον Οκτώβριο του 2001 και η πιο πρόσφατη έκδοση (**Nitf 3.2**), που κυκλοφόρησε τον Οκτώβριο του 2003.

➡ **XMLNews**

Το πρότυπο **XMLNews** αποτελεί ίσως ένα από τα πιο ανεπτυγμένα μορφότυπα που χρησιμοποιούνταν από τη βιομηχανία ειδήσεων στον Παγκόσμιο Ιστό μέχρι πρόσφατα. Σχεδιάστηκε από τον David Megginson (2000) και είναι υποδιέστερο του **Nitf**. Το κομμάτι της περιγραφής είναι γνωστό ως «**XMLNews-Story**». Επιπροσθέτως, το **XMLNews** περιείχε το λεγόμενο «**XMLNews-Meta**», μία

εφαρμογή σε **RDF** (Resource Description Framework) για την περιγραφή του περιεχομένου των ειδήσεων.

Το **XMLNews-Meta** είναι ένα επεκτάσιμο λεξιλόγιο που χρησιμοποιείται για την περιγραφή ειδησεογραφικών πηγών. Τα βασικά χαρακτηριστικά του συνίστανται στην ικανότητά του να περιγράφει τα εξής ακόλουθα:

- ◆ **Ταύτιση (Identification)**: προσδιορίζει ένα μοναδικό ID στην πηγή η οποία περιγράφεται
- ◆ **Πληροφορίες επικεφαλίδας (Header Information)**: παρέχει πληροφορίες για τη γλώσσα, τον τίτλο και γενικότερα για την περιγραφή του ειδησεογραφικού αντικειμένου
- ◆ **Ορόσημα (Milestones)**: αναφέρεται σε στοιχεία όπως, η έκδοση, η κυκλοφορία, η λήψη ειδήσεων και οι χρόνοι λήξης
- ◆ **Προέλευση (Provenance)**: δηλώνεται η προέλευση – πηγή της ειδησεογραφικής ιστορίας
- ◆ **Πνευματικά δικαιώματα (Rights)**: εδώ έχουμε πληροφορίες σχετικά με τον κάτοχο πνευματικών δικαιωμάτων (copyright) και τα δικαιώματα διάθεσης
- ◆ **Θεματικό υλικό (Subject Matter)**: παρέχονται πληροφορίες σχετικά με την αυτόματη ταξινόμηση
- ◆ **Σύνδεση (Linking)**: περιγράφονται οι εσωτερικές συσχετίσεις μέσα στην ειδησεογραφική ιστορία, π.χ. σύγχρονη έκδοση σε σχέση με παλιές εκδόσεις

Επιπροσθέτως, το **XMLNews-Meta** εκφράζεται σε σχήμα RDF και είναι επεκτάσιμο με συνέπεια να καθίσταται δυνατή η χρήση του από οργανισμούς, που θέλουν να το παραμετροποιήσουν και να το επεκτείνουν με βάση τις ανάγκες και τους σκοπούς τους.

➡ **PRISM**

Το πρότυπο **PRISM** (Publishing Requirements for Industry Standard Metadata), το οποίο δημιουργήθηκε και λειτουργήσε υπό την αιγίδα της εταιρείας **IDEAlliance** (International Digital Enterprise Alliance), καθορίζει ένα λεξιλόγιο μεταδεδομένων βασισμένο στη γλώσσα **XML** και χρησιμοποιείται για τη διαχείριση, τη συγκομιδή, τη μετεξέλιξη και την πολυσκοπιμότητα περιοδικών, ειδήσεων, καταλόγων, βιβλίων και το δεσπόζον περιεχόμενό τους. Το πεδίο ενδιαφέροντος για τη δημιουργία του προτύπου **PRISM** ξεκινά από τις ανάγκες των εκδοτών να λαμβάνουν, να παρακολουθούν και να μεταφέρουν πολυτμηματικό περιεχόμενο. Το συγκεκριμένο πρότυπο εστιάζει στις πρόσθετες χρήσεις του περιεχομένου και τα μεταδεδομένα, που αφορούν την εμφάνισή του βρίσκονται έξω από τα ενδιαφέροντα του.

Επιπλέον, οι ιθύνοντες της βιομηχανίας ειδήσεων κατάλαβαν πολύ σύντομα ότι η παρακολούθηση του περιεχομένου μέσα από τον κύκλο ζωής του - τη δημιουργία, τη χρήση, το σύνολο, την κοινοπραξία και την επαναχρησιμοποίηση των ψηφιακών στοιχείων – θα ήταν μία δαπανηρή εκ των προτέρων προσπάθεια χωρίς την οργάνωσή του με τη χρήση προτύπων. Έπρεπε, λοιπόν, να καταλήξουν σε μία συμφωνία σχετικά με τα λεξιλόγια και τους ορισμούς που καθορίζουν πως ένα στοιχείο περιγράφεται, ανταλλάσσεται και εκτίθεται, έτσι ώστε με αυτόματο τρόπο να αναπτυχθούν και νέα εργαλεία και συγκεκριμένες διαδικασίες.

Το **Prism** συστήνει τη χρήση συγκεκριμένων υπαρχόντων προτύπων, μερικά των οποίων είναι το **DC** (Dublin Core), το **RDF**, το **NewsML**, το **NITF**, το **RSS** (Really Simple Syndication). Επίσης, προτείνει διάφορες προδιαγραφές του *ISO* για τοποθεσίες, γλώσσες, και μορφότυπα ημερομηνιών και ωρών. Στις μέρες μας, το

PRISM αποτελείται από δύο προδιαγραφές. Η μία είναι το ίδιο το πρότυπο **PRISM** το οποίο παρέχει τον ορισμό για όλο το πλαίσιο των εργασιών του και η δεύτερη είναι το **PRISM Aggregator DTD** που αποτελεί ένα καινούργιο μορφότυπο για τους εκδότες. Το πρότυπο **PRISM** έχει τρεις εκδόσεις. Η πρώτη έκδοση (**PRISM 1.0**) πραγματοποιήθηκε τον Απρίλιο του 2001, ενώ η δεύτερη (**PRISM 1.1**) κυκλοφόρησε ένα χρόνο αργότερα (2002). Τέλος η τρίτη του έκδοση (**PRISM 1.2**) κυκλοφόρησε τον Αύγουστο του 2004 κατά την οποία η Ομάδα Εργασίας του **PRISM** ανέπτυξε ένα πρότυπο σε μορφή XML, το οποίο συνδυάζει τα μεταδεδομένα του **PRISM** μαζί με ενσωματωμένο περιεχόμενο, προκειμένου να υποστηρίξει αυτόματη μετάδοση περιεχομένου από τους εκδότες στους διανομείς και συλλέκτες πληροφοριών.

➡ **ICE**

Η ανακοίνωση της δημιουργίας του πρωτοκόλλου **ICE** (Information and Content Exchange) τον Οκτώβριο του 1998 από την εταιρεία **IDEAlliance** έδωσε μία νέα ώθηση όσον αφορά το αντικείμενο «Ανταλλαγή Περιεχομένου και Πληροφορίας». Η δημιουργία και ολοκλήρωση του πρωτοκόλλου που βασιζόταν στην γλώσσα σήμανσης XML και προοριζόταν για χρήση στον Παγκόσμιο Ιστό αποτέλεσε ένα σημαντικό βήμα για την υποστήριξη της διανομής πληροφοριών μέσα στην εκδοτική και όχι μόνο κοινωνία. Ήταν στην ουσία ένα πρωτόκολλο, το οποίο αφορούσε την επόμενη γενιά του Παγκόσμιου Ιστού.

Η βασική αποστολή του **ICE** είναι να διευκολύνει την ελεγχόμενη διαχείριση και ανταλλαγή ηλεκτρονικών μονάδων μεταξύ των ενδιαφερόμενων συνεργατών και να συνδέεται με τον Παγκόσμιο Ιστό. Οι εφαρμογές, που βασίζονται στο συγκεκριμένο πρωτόκολλο, επιτρέπουν στις εταιρείες να κατασκευάζουν με σχετικά εύκολο τρόπο εκδοτικά δίκτυα διανομής καθώς υιοθετούν και χρησιμοποιούν υπηρεσίες του Παγκόσμιου Ιστού οι οποίες στηρίζονται σε ποικίλα δίκτυα πληροφορίας. (Hammond, Hannay and Lund, 2004, p. 2)

Το **ICE** παρέχει μία δομή και μία αρχιτεκτονική που υποστηρίζουν την αυτόματη ενημέρωση, επεξεργασία και έλεγχο των αντικειμένων με έναν αξιόπιστο τρόπο χωρίς χειρωνακτικό πακετάρισμα ή γνώση της δομής άλλων ιστοσελίδων. Όσον αφορά τις ιστοσελίδες με τη μεγαλύτερη ζήτηση, οι τελικοί χρήστες επωφελούνται από προορισμούς του Ιστού τους περισσότερο ολοκληρωμένους και εύκολους σε χρήση, οι οποίοι μειώνουν το ενδεχόμενο της αποτυχίας να πρέπει να ψάχνουν οι χρήστες μέσα σε ανεπαρκείς και στενά εστιασμένες ιστοσελίδες για να μπορέσουν να βρουν αυτό που πραγματικά χρειάζονται. Το πρωτόκολλο **ICE** έχει δύο εκδόσεις σχετικά με τις τεχνικές προδιαγραφές του. Η πρώτη (**ICE 1.0**) εκδόθηκε και κυκλοφόρησε το 1998, ενώ τέθηκε ως Σημείωμα στο **W3C** (World Wide Web Consortium) το 1999. Η δεύτερη (**ICE 2.0**) εκδόθηκε τον Δεκέμβριο του 2003, ενώ ως τελική έκδοση (version) ανακοινώθηκε το Αύγουστο του 2004.

➡ **IPTC 7901**

Το **IPTC** (International Press Telecommunications Council) **7901** αναπτύχθηκε από τον οργανισμό **IPTC**, χρησιμοποιούνταν για τη μετάδοση μηνυμάτων κειμένου σε εφημερίδες, πρακτορεία ειδήσεων και άλλους παραλήπτες ειδήσεων. Η πρώτη του έκδοση εμφανίστηκε στις αρχές της δεκαετίας του 1980 και ενημερωνόταν σε τακτά χρονικά διαστήματα μέχρι και την τελευταία του έκδοση (**IPTC 7901 5.0**), η οποία εγκρίθηκε το 1995. Από τότε, όμως, και μετά η ανάπτυξη του **IPTC 7901** σταμάτησε παρόλο που χρησιμοποιείται σε πολλές χώρες ακόμη και σήμερα.

Προοριζόταν για διεθνή χρήση και για αυτό το λόγο υπολόγιζε τεχνικές και γλωσσικές διαφορές ανάμεσα σε διάφορες χώρες και σχεδιάστηκε με γνώμονα τη χρήση του σε πολυάριθμες γλώσσες και αλφάβητα.

▼ **IIIM**

Το **IIIM** (Information Interchange Model) υπήρξε και το πρώτο μορφότυπο ανταλλαγής πολυμεσικών ειδήσεων. Σχεδιάστηκε και αναπτύχθηκε μέσα από τη συνεργασία του **IPTC** και της Αμερικανικής Ένωσης Εφημερίδων / **NAA**. Οι συγκεκριμένοι οργανισμοί είχαν ως σκοπό τη δημιουργία ενός εφαρμόσιμου σε παγκόσμιο επίπεδο μοντέλου για τη διαχείριση όλων των ειδών των δεδομένων. Έτσι, λοιπόν, προέκυψε το 1991 μέσα από αυτή την προσπάθεια η πρώτη έκδοση του **IIIM** (**IIIM 1.0**), η οποία και στη συνέχεια δέχθηκε αρκετές βελτιώσεις. Η ανάπτυξη των νέων τεχνολογιών σε συνδυασμό με την εμφάνιση της XML όσον αφορά την επανεμφάνιση των δεδομένων οδήγησε στο πάγωμα της περαιτέρω ανάπτυξης του συγκεκριμένου προτύπου το 1997. Η πιο πρόσφατη και τελευταία έκδοση του **IIIM** είναι η **IIIM 4.1**.

Τα στοιχεία των μεταδεδομένων του είναι γνωστά ως «**IPTC headers**». Εταιρείες, όπως η Adobe Systems Inc. έχουν εφεύρει έναν δικό τους μηχανισμό για να εισάγουν δομές μεταδεδομένων μέσα στο Photoshop, JPEG και TIFF αρχεία, και έχουν υιοθετήσει τη δομή των δεδομένων του **IIIM** καθώς και αρκετά από τα στοιχεία των μεταδεδομένων του. Ο μηχανισμός εισαγωγής μεταδεδομένων εφαρμόστηκε επίσης και από άλλους παραγωγούς λογισμικού και με αυτόν τον τρόπο πολλά προγράμματα βιβλιοθηκών είναι ικανά να γράφουν και να διαβάζουν «**IPTC headers**».

Πέρα από όλα αυτά, το πρότυπο **IIIM** σχεδιάστηκε με σκοπό την παροχή επικοινωνίας σε παγκόσμιο επίπεδο και τη διαχείριση όλων των ειδών δεδομένων, όπως κείμενο, φωτογραφίες, γραφικά, κλπ., μέσα σε ένα μοναδικό δίκτυο ή ένα μοναδικό μέσο αποθήκευσης. Επιπροσθέτως, παρέχεται ένας μηχανισμός για να χρησιμοποιεί υπάρχοντα μορφότυπα κατά τη διάρκεια της μετάδοσης των δεδομένων.

Κωδικοποίηση με το Πρότυπο NewsML

Πληροφορίες σχετικά με το Ειδησεογραφικό Αντικείμενο

Στα πλαίσια της συγκεκριμένης εισήγησης κρίθηκε αναγκαία και εξαιρετικά χρήσιμη η εφαρμογή και η υλοποίηση της κωδικοποίησης ενός ειδησεογραφικού αντικειμένου, το οποίο αποτελείται από μία σειρά επιμέρους αντικειμένων, που συνθέτουν όλα μαζί ένα ειδησεογραφικό αντικείμενο με σχετικό θεματικό περιεχόμενο. Θα πραγματοποιηθεί η κωδικοποίηση τριών κειμένων και μίας εικόνας με βάση το πρότυπο **NewsML**. Το ειδησεογραφικό αντικείμενο, το οποίο δημιουργείται για να αποτελέσει και μία είδηση στο σύνολό της με σκοπό την αποστολή της σε μία άλλη εφημερίδα, συνίσταται στα εξής μέρη:

- ❖ Ένα βιογραφικό άρθρο για τη ζωή και το έργο του Κωνσταντίνου Π. Καβάφη, που προέρχεται από την εφημερίδα «ΕΛΕΥΘΕΡΟΝ ΒΗΜΑ» της χρονικής περιόδου 1922 – 1944
- ❖ Μία εικόνα / σχέδιο του Κωνσταντίνου Π. Καβάφη, που συνοδεύει το συγκεκριμένο άρθρο

- ❖ Ένα ποίημα του Κωνσταντίνου Π. Καβάφη στην ελληνική γλώσσα με τίτλο «Περιμένοντας τους βαρβάρους», το οποίο παρατίθεται ολόκληρο μέσα στο βιογραφικό άρθρο
- ❖ Η έκδοση του ποιήματος στην αγγλική γλώσσα με τίτλο «Waiting for the Barbarians», δοσμένη από την ποιήτρια Dana Bonstrom σε μετάφραση του Στρατή Χαβιάρα στις Εκδόσεις «Ερμής» (2004), στο βιβλίο με τίτλο «C. P. CAVAFY, the Canon. The Original One Hundred and Fifty – four Poems»

Η συγκεκριμένη κωδικοποίηση, η οποία προτείνεται και θα παρατεθεί ακολούθως, αποτελεί ένα έργο σύνθετο. Σκοπός της συγκεκριμένης προσπάθειας είναι να αναδείξει μία νέα πρόταση όσον αφορά τη σύνθεση ενός ειδησεογραφικού αντικειμένου, το οποίο συνίσταται σε επιμέρους αντικείμενα διαφορετικού είδους, προερχόμενα από διαφορετικές πηγές και αποδιδόμενα σε διαφορετικές γλώσσες.

Η συντακτική και εννοιολογική περιγραφή αυτών των αντικειμένων καθώς επίσης και η δυνατότητα ομαδοποίησης και απόδοσής τους από το πρότυπο **NewsML** αποτελεί ένα ενθαρρυντικό και ουσιαστικό βήμα. Στην ουσία τέσσερα διαφορετικά αντικείμενα μπορούν να συγκροτήσουν κάτι ενιαίο, όταν έχουν ένα κοινό χαρακτηριστικό γνώρισμα. Βασίζονται σε μία ενιαία θεματική ενότητα, η οποία αντικατοπτρίζεται μέσα από την περιγραφή και την ενοποίηση των παραπάνω ειδησεογραφικών αντικειμένων.

Η υλοποίηση αυτή αποτελεί παράλληλα και ένα σημαντικό εγχείρημα για την κοινωνία της πληροφόρησης, γιατί αναδεικνύει τις απεριόριστες δυνατότητες και συνεργασίες που μπορεί να προκύψουν μέσα από αυτή την διαδικασία. Τα διάφορα είδη μεταδεδωμένων που ενσωματώνονται στο ειδησεογραφικό τεκμήριο κατά την διαδικασία της κωδικοποίησης καθώς και τα δεδομένα εκείνα που αφορούν πληροφορίες σχετικά με τη μετάδοση της είδησης συντελούν στην πραγματοποίηση του στόχου της όλης προσπάθειας που είναι η άμεση, εύστοχη, και ουσιώδης πληροφόρηση.

Επιλογή του προτύπου & ετικέτες περιγραφής του Ειδησεογραφικού Αντικειμένου

Η επιλογή του συγκεκριμένου προτύπου (**NewsML**) για την ακολουθούμενη περιγραφή και παράσταση των δεδομένων έγινε σύμφωνα με τέσσερα σημαντικά κριτήρια:

- ◆ Το είδος των δεδομένων που έπρεπε να κωδικοποιηθούν
- ◆ Το επίπεδο της κωδικοποίησης
- ◆ Την επικράτηση, αναγνώριση και υιοθέτηση του συγκεκριμένου προτύπου στον χώρο της ειδησεογραφικής εκδοτικής κοινωνίας
- ◆ Τις ανάγκες περιγραφής του ειδησεογραφικού αντικειμένου και τις απαιτήσεις των χρηστών και διαχειριστών αυτού

Η περιγραφή του ειδησεογραφικού αντικειμένου ξεκινά με την xml έκδοση 1.0 σε κωδικοποίηση UTF-8 (<?xml version="1.0" encoding="UTF-8"?>). Στη συνέχεια ακολουθεί η δήλωση του DOCTYPE (<!DOCTYPE NewsML SYSTEM "NewsML_1.2.dtd">) και τέλος δίνεται η κωδικοποίηση του **NewsML** ειδησεογραφικού τεκμηρίου, το οποίο συνίσταται σε τέσσερα ξεχωριστά **NewsItems**. Η επιλογή των ετικετών για την περιγραφή των δεδομένων πραγματοποιήθηκε με βάση τα στοιχεία, που έπρεπε να κωδικοποιηθούν για τα συγκεκριμένα **NewsItems** και βοήθησε ουσιαστικά και σημαντικά στην όλη διαδικασία της υλοποίησης.

Επιλογή XML EDITOR για την κωδικοποίηση του αντικειμένου

Όσον αφορά την κωδικοποίηση του ειδησεογραφικού αντικειμένου επιλέχθηκε, ύστερα από μελέτη, ο **XML Spy 2005 Home Edition**.

Το **NewsML DTD** εισήχθη στον *XML EDITOR* και ένα καινούργιο έγγραφο δημιουργήθηκε, για να επιλεχθούν στη συνέχεια τα στοιχεία (**elements**) σύμφωνα με τα οποία πραγματοποιήθηκε η περιγραφή του ειδησεογραφικού αντικειμένου (**NewsObject**). Σε κάθε βήμα, αλλά και με την ολοκλήρωση της διαδικασίας της κωδικοποίησης έγινε έλεγχος εγκυρότητας (**validation**) της δομής του XML τεκμηρίου.

Διαδικασία μετατροπής της XML μορφής του Ειδησεογραφικού Αντικειμένου σε HTML μορφή

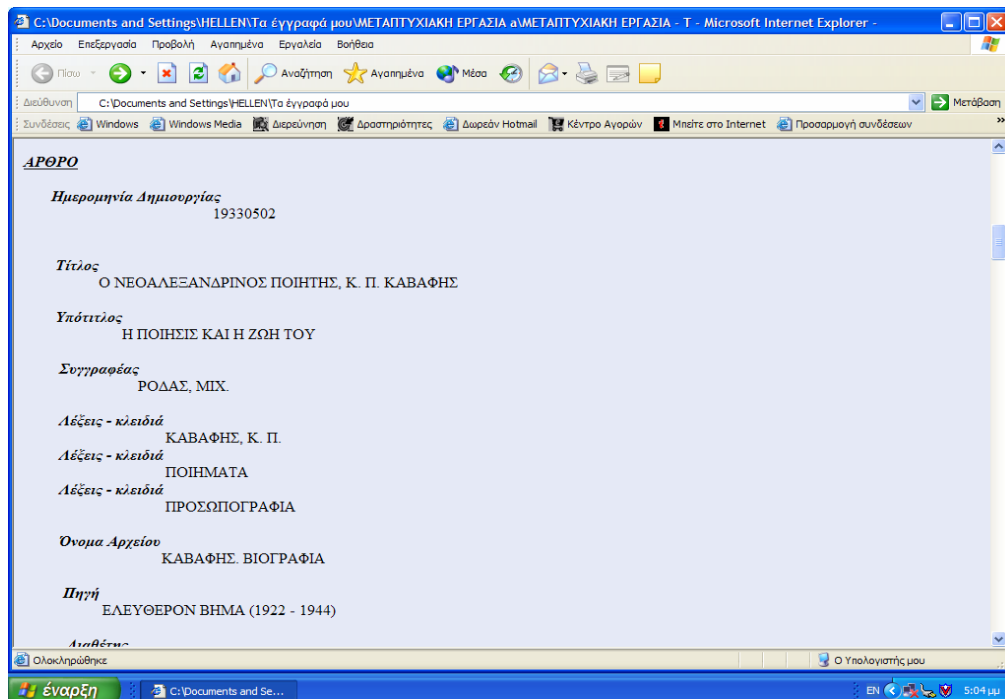
Η ακριβής διαδικασία, η οποία ακολουθήθηκε για να μετατραπούν και να απεικονισθούν τα βιβλιογραφικά δεδομένα από την XML μορφή σε μία σελίδα του Ιστού με την μορφή HTML είναι ακολούθως η παρακάτω :

- 1) Κατ' αρχήν εντοπίσθηκε το DTD, το οποίο επικυρώνει δεδομένα. Πρόκειται για το **NewsML_1.2.dtd**. (http://www.newsml.org/pages/spec_main.php).
- 2) Τοποθετήθηκε το συγκεκριμένο DTD στην εφαρμογή Stylevision της Altova, η οποία ανέλαβε να διαβάσει το σχήμα του DTD. (<http://www.altova.com>)
- 3) Επιπροσθέτως, εισήχθη το XML έγγραφο **XMLData.xml** στην εφαρμογή Stylevision.

```
<?xml version="1.0" encoding="utf-8" ?>
<!-- <?xml-stylesheet type="text/xsl" href="genhtml.xslt" ?> -->
<!DOCTYPE NewsML (View source for full doctype...)>
- <NewsML>
- <NewsEnvelope>
- <TransmissionId>lpa.EB.3964.ΚΑΒΑΦΗΣ</TransmissionId>
- <SentFrom>
- <Party FormalName="ΕΛΛΗΝΙΚΟ ΨΗΦΙΑΚΟ ΑΡΧΕΙΟ ΤΥΠΟΥ ΚΑΙ ΜΜΕ 2005:Δ.Ο.Α." />
</SentFrom>
- <DateAndTime>20050228</DateAndTime>
</NewsEnvelope>
- <NewsItem>
- <Identification>
- <NewsIdentifier>
- <ProviderId>lpa.gr</ProviderId>
- <DateId>20050228</DateId>
- <NewsItemId>lpa.EB.3964</NewsItemId>
- <RevisionId PreviousRevision="0" Update="1" />
- <PublicIdentifier>urn:newsml:lpa.gr:20050228:lpa.EB.3964.ΚΑΒΑΦΗΣ:1</PublicIdentifier>
</NewsIdentifier>
</Identification>
- <NewsManagement>
- <NewsItemType FormalName="ΑΡΘΡΟ" />
- <FirstCreated>19330502</FirstCreated>
- <ThisRevisionCreated>20010625</ThisRevisionCreated>
- <Status FormalName="usable" />
</NewsManagement>
- <NewsComponent Essential="no" EquivalentsList="no">
- <NewsLines>
- <HeadLine>Ο ΝΕΟΑΛΕΞΑΝΔΡΙΝΟΣ ΠΟΙΗΤΗΣ, Κ. Π. ΚΑΒΑΦΗΣ</HeadLine>
- <SubHeadLine>Η ΠΟΙΗΣΙΣ ΚΑΙ Η ΖΩΗ ΤΟΥ</SubHeadLine>
- <ByLine>ΡΟΔΑΣ, ΜΙΧ.</ByLine>
- <CopyrightLine>ΕΛΛΗΝΙΚΟ ΨΗΦΙΑΚΟ ΑΡΧΕΙΟ ΤΥΠΟΥ ΚΑΙ ΜΜΕ 2005:Δ.Ο.Α.</CopyrightLine>
- <KeywordLine>ΚΑΒΑΦΗΣ, Κ. Π.</KeywordLine>
- <KeywordLine>ΠΟΙΗΜΑΤΑ</KeywordLine>
- <KeywordLine>ΠΡΟΣΩΠΟΓΡΑΦΙΑ</KeywordLine>
</NewsLines>
- <AdministrativeMetadata>
- <FileName>ΚΑΒΑΦΗΣ. ΒΙΟΓΡΑΦΙΑ</FileName>
- <Provider>
- <Party FormalName="ΕΛΛΗΝΙΚΟ ΨΗΦΙΑΚΟ ΑΡΧΕΙΟ ΤΥΠΟΥ ΚΑΙ ΜΜΕ 2005:Δ.Ο.Α." />
</Provider>
- <Creator>
- <Party FormalName="ELENI" />
</Creator>
- <Source>
```

4)Το περιβάλλον της εφαρμογής Stylevision, μέσω ενός editor τον οποίο διαθέτει, έδωσε την δυνατότητα να παραχθεί το αρχείο **genhtml.xslt**, χρησιμοποιώντας ως δεδομένα εισαγωγής τα δύο παραπάνω αρχεία (NewsML_1.2.dtd & XMLData.xml).

5)Ο ρόλος του αρχείου genhtml.xslt είναι να μετατρέψει και να απεικονίσει το XML έγγραφο σε μορφή HTML.



Συμπεράσματα

Η επισκόπηση των διαφόρων προτύπων και εργαλείων για τη διαχείριση ειδήσεων απέδειξε ότι επικρατούν στο χώρο της κοινωνίας της πληροφόρησης δυνατότητες και τεχνολογίες, που έχουν άμεσο στόχο την πραγματοποίηση της σωστής οργάνωσης, τεκμηρίωσης και διαχείρισης των πληροφοριών.

Οι πληροφορίες πρέπει να δομούνται με ένα συγκεκριμένο τρόπο, έτσι ώστε να αποκτούν νόημα οι έννοιες και οι σημασίες που εκφράζουν. Εδώ έρχονται να κάνουν αισθητή την παρουσία τους τα πρότυπα, που σαν σκοπό έχουν να περιγράψουν, να δομήσουν τις πληροφορίες και να βοηθήσουν ουσιαστικά όχι μόνο στην τεκμηρίωση των ειδησεογραφικών πληροφοριών, αλλά και στην ανταλλαγή τους μέσω της χρήσης τεχνολογιών, οι οποίες επιτρέπουν στα διαφορετικά συστήματα να τις καταλαβαίνουν.

Συνεπώς, σημαντικό βήμα αποτελεί και η διαλειτουργικότητα των συστημάτων, που κατέχουν την όποια πληροφορία. Το **NewsML** αποτελεί ένα από τα πρότυπα και η γλώσσα σήμανσης **XML** μία τεχνολογία. Αυτό το πρότυπο, όπως το γνωρίσαμε, αντικατοπτρίζει την αυξανόμενη σημαντικότητα των ειδησεογραφικών ιστοριών και αποτελεί ένα πρότυπο, που δημιουργεί και προβάλλει την αξία της είδησης.

Επιπροσθέτως, είναι ένα εντελώς καινούργιο πρότυπο, το οποίο συνεχώς επιδέχεται προσθήκες και επεκτάσεις με σκοπό την πληρέστερη και καλύτερη εφαρμογή του. Αυτό διαφαίνεται και από το γεγονός ότι κάποιοι συγκεκριμένοι οργανισμοί (Πρακτορείο Ρούτερ, Γαλλικό Πρακτορείο Ειδήσεων), που το χρησιμοποιούν, παρέχουν τη βοήθειά τους για την περαιτέρω ανάπτυξή του, ενώ παράλληλα ο οργανισμός IPTC, που αποτελεί και παραγωγό του, κάνει προσπάθειες για τη βελτίωση του και τη συνεργασία του με άλλα πρότυπα περιγραφής βιβλιογραφικών δεδομένων.

Σε μία κοινωνία – όπως η σημερινή χαρακτηριζόμενη και ως Κοινωνία της Πληροφορίας – μέσα στην οποία καθημερινά οι κάθε είδους πληροφορίες προκύπτουν με γρήγορο και άμεσο τρόπο, η χρήση των προτύπων και των διαφόρων τεχνολογιών για την οργάνωση και διάθεσή τους αποτελεί μία επιτακτική και συγχρόνως ουσιαστική ανάγκη.

Βιβλιογραφία

1. Εκπαιδευτική Πύλη του Υ.Π.Ε.Π.Θ. (2007) **Κείμενα για τον Κ. Π. ΚΑΒΑΦΗ**. Available from: <<http://www.e-yliko.gr/Fyyl/glossa/fikavkrit1.htm>> [Accessed 02 July 2007]
2. Σπουδαστήριο Νέου Ελληνισμού (2007) **Αρχείο ΚΑΒΑΦΗ: Εργογραφία – Αναγνωρισμένα Ποιήματα**. Available from: <<http://www.snhell.gr/anagnwr.html>> [Accessed 02 July 2007]
3. Bide, M. & Associates (2000) **Standards for Electronic Publishing”, an overview, a report for the NEDLIB project**, pp. 12-13. Available from: <<http://www.kb.nl/coop/nedlib/results/e-publishingstandards.pdf>>, [Accessed 20 July 2007].
4. Dumbill, Edd (2000) **XML in News Syndication**. Available from: <<http://www.xml.com>> [Accessed 28 June 2007].
5. Goodman, Hays (2002) **Number of dedicated newspaper archive solutions falls as integration grows**. The International Journal of Newspaper Technology. Available from: <http://www.newsandtech.com/issues/2002/06-02/ot/06-02_archives.htm> [Accessed 10 August 2007]
6. Hammond, Tony, Hannay, Timo and Ben Lund (2004) **The Role of RSS in Science Publishing: Syndication and Annotation on the Web**, D-Lib Magazine, 10 (12). Available from: <<http://www.dlib.org/dlib/december04/hammond/12hammond.html>> [Accessed 05 August 2007]
7. NewsML.org (2004) **NewsML Guidelines**. Available from: <http://www.newsml.org/IPTC/NewsML/1.2/documentation/NewsML_1.2-doc-Guidelines_1.00pdf> [Accessed 31 July 2007]
8. Rabin, Jo (2004) **News Standards: A Rising Tide of Commoditization** Available from: <<http://www.xml.com/pub/a/2004/05/05/news.html>> [Accessed 01 August 2007]